Merging Real-Time Channel Sensor Networks with Continental-Scale Hydrologic Models: A Data Assimilation Approach for Improving Accuracy in Flood Depth Predictions

Amir Javaheri 1, Mohammad Nabatian 2, Ehsan Omranian 3,*, Meghna Babbar-Sebens 1 and Seong Jin Noh 2

1 Department of Civil and Construction Engineering, Oregon State University, Corvallis, OR 97331, USA; javaheam@oregonstate.edu (A.J.); meghna@oregonstate.edu (M.B.-S.)
2 Department of Civil Engineering, University of Texas at Arlington, Arlington, TX 76019, USA; mohammad.nabatian@mavs.uta.edu (M.N.); seongjin.noh@uta.edu (S.J.N.)
3 Department of Civil & Environmental Engineering, University of Texas at San Antonio, San Antonio, TX 78256, USA

* Correspondence: seyedehsan.omranian@utsa.edu; Tel.: +1-210-803-3847

Received: 27 October 2017; Accepted: 18 January 2018; Published: 21 January 2018

Abstract: This study proposes a framework that (i) uses data assimilation as a post processing technique to increase the accuracy of water depth prediction, (ii) updates streamflow generated by the National Water Model (NWM), and (iii) proposes a scope for updating the initial condition of continental-scale hydrologic models. Predicted flows by the NWM for each stream were converted to the water depth using the Height Above Nearest Drainage (HAND) method. The water level measurements from the Iowa Flood Inundation System (a test bed sensor network in this study) were converted to water depths and then assimilated into the HAND model using the ensemble Kalman filter (EnKF). The results showed that after assimilating the water depth using the EnKF, for a flood event during 2015, the normalized root mean square error was reduced by 0.50 m (51%) for training tributaries. Comparison of the updated modeled water stage values with observations at testing locations showed that the proposed methodology was also effective on the tributaries with no observations. The overall error reduced from 0.89 m to 0.44 m for testing tributaries. The updated depths were then converted to streamflow using rating curves generated by the HAND model. The error between updated flows and observations at United States Geological Survey (USGS) station at Squaw Creek decreased by 35%. For future work, updated streamflows could also be used to dynamically update initial conditions in the continental-scale National Water Model.

Keywords: data assimilation; ensemble Kalman filter; flood inundation maps; National Water Model (NWM)

1. Introduction

Flooding is among the most destructive natural disasters globally. In the United States, based on a U.S. National Weather Service (NWS) report, the average annual property/human losses are estimated to be more than $8 billion [1–3] (Federal Emergency Management Agency (FEMA), 2013). Flood inundation maps help to detect flood-prone regions and can prevent major catastrophe by providing reliable information to the public about the flood-risk [4,5]. However, these maps are based on predictions of water depth values in the stream and on the landscape for extreme rainfall events [6]. Hydrology and hydraulic models can be useful for predicting these depth values, although, like many other models, these models are only as reliable as the underlying assumptions in the model’s structure.
and parameters [7,8]. The assimilation of water depth measurements, as a post-processing technique, has the potential to reduce the error between model predictions and observations in order to generate more reliable flood inundation maps [9].

Sequential data assimilation techniques are often used for updating a model’s state variables when new observations become available [10]. A data assimilation process is also able to reduce the uncertainty in prediction by integrating real-time observations from a variety of monitoring technologies [11,12]. The Kalman filter [13] is a commonly used data assimilation technique that was initially developed to update the state variables of linear systems [14]. However, this method has been used for nonlinear problems as well [15]. The ensemble Kalman filter (EnKF) is another data assimilation technique that was introduced by Evensen [16]. In the case of non-linear models for which the assumption of linearity is not satisfied, EnKF can be used as an effective technique. Multiple studies have also used this method in the past to update hydraulic, hydrologic, and hydrodynamic models [17–19].

The main objectives of this study include:

1. Assimilating water depth measurements to dynamically update water level predictions.
2. Improving streamflow predicted by the National Water Model (NWM) using updated water levels.
3. Proposing a scope to update continental-scale hydrologic models (e.g., NWM).

2. Study Area and Methodology

2.1. Study Area Characteristics and Data Collection

The study domain is the Squaw Creek watershed (Hydrologic Unit Code (HUC) ID = 0708010503), located in Jasper County, Iowa (Figure 1), which drains 40.3 km² and discharges into the Skunk River. It is located on Southern Iowa Drift Plain and characterized by steeply rolling hills and well developed drainage [20]. This watershed has experienced several flooding events in the past [21], and it contains a dense network of bridge sensors that measure the water level along the channel. Flow measurements are collected at a USGS station (# 5470500) in this watershed. Water surface elevation were gathered from the Iowa Flood Information System (IFIS). The Iowa Flood Center (IFC) developed and maintains nearly 250 stream stage sensors across the state. Sensors are attached to the sides of bridges and are designed to measure the distance from the sensor to the water surface. Data is transmitted automatically and frequently to the Iowa Flood Information System. Water surface elevation at each time step can be calculated by subtracting the sensor measurements from the elevation of each sensor.

Figure 1. Squaw Creek watershed (Iowa State).
2.2. Proposed Approach

Figure 2 illustrates the overall proposed methodology:

Hourly predicted flows for each tributary were first obtained from the NWM (Step 1). The Height Above Nearest Drainage (HAND) [22] method was then used to predict the water depth in river channels (Step 2). Incoming hourly water level observations from the Iowa Flood Information System (IFIS) were converted to water depth (Step 3) to become consistent with HAND water depth predictions, and then assimilated into the HAND model via an ensemble Kalman filter technique (Step 4). Finally, using back calculation, updated water levels were converted to flow using rating curves generated by the HAND (Step 5). For future work, results from Steps 4 and 5 can be used to create flood inundation maps and update the initial conditions of the hydrologic model (NWM), respectively. Finally, updated streamflows were validated using flow measurements at USGS stations (Step 6).

There are 85 tributaries and 20 bridge sensors that measure the water levels. Since the cross-sections were not available for all of these sensor locations, measurements of only 10 sensors were used for data assimilation. Six sensors were randomly selected for training and the other four sensors were selected for testing to make sure that this approach is also suitable for tributaries with no observation. The NWM predicts the flow at the outlet of each sub-basin, however, bridge sensors are located upstream of outlets. Assuming equal soil type and land use for each sub-basin, the flow at the outlet was linearly distributed along the river.

\[
Q_{outlet} = \frac{L_{Total}}{L_{Total} - L_{Partial}} \times Q_B
\]

where \(Q_{outlet}\) is the flow at the outlet of each sub-basin, \(L_{Total}\) is the length of river, \(L_{Partial}\) is the river length from bridge sensor to the outlet of each sub-basin, and \(Q_B\) is the flow measured at each bridge sensor.
2.3. National Water Model

The National Water Model (MWM) is a continental-scale hydrologic model that generates forecasts for multiple variables [23]. The NWM was released by the National Weather Service (NWS) Office of Water Prediction (OWP) in collaboration with the National Center for Atmospheric Research (NCAR) and the National Center for Environmental Prediction (NCEP). The model generates streamflows for 2.7 million reaches of the National Hydrography Dataset (NHD) with four forecast products (i.e., analysis and assimilation, short range forecast, medium range forecast, and long range forecast) [24]. Table 1 shows the forecast step, frequency, and forecast duration of each product.

Table 1. National Water Model details (http://water.noaa.gov/about/nwm).

<table>
<thead>
<tr>
<th>Forecast Product</th>
<th>Forecast Latency</th>
<th>Frequency</th>
<th>Forecast Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis and assimilation</td>
<td>1 h</td>
<td>Hourly</td>
<td>0–3 h</td>
</tr>
<tr>
<td>Short range forecast</td>
<td>1 h</td>
<td>Hourly</td>
<td>0–15 h</td>
</tr>
<tr>
<td>Medium range forecast</td>
<td>3 h</td>
<td>Daily</td>
<td>0–10 days</td>
</tr>
<tr>
<td>Long range forecast</td>
<td>6 h</td>
<td>4× Daily</td>
<td>0–30 days</td>
</tr>
</tbody>
</table>

2.4. The HAND Method

The Height Above Nearest Drainage (HAND) [22] is a model generated by the National Science Foundation (NSF) and the National Water Center for developing continental-scale inundation mapping [25]. It is based on hydrological terrain analysis to produce flood inundation maps [26,27]. The Digital Elevation Model (DEM) of watershed is used to drive the flow direction and flow accumulation area. A stream grid is then generated to be used as an input to Terrain Analysis Using Digital Elevation Models [28], along with the flow-direction grid, to create the height of each grid cell above the nearest drainage. The NWM predicts flow at the outlet, which can be converted to water depth with a rating curve, and 10 sensors also provide information of water depth. The HAND is a raster-based method and specifies the inundation zone according to the corresponding river segment. It produces an inundation map according to the water depth of tributaries [25].

2.5. Ensemble Kalman Filter

The ensemble Kalman filter algorithm was used in the proposed methodology (Step 4) to assimilate water depth observations into the model. In the ensemble Kalman filter, the prediction model is represented by:

\[ h_k = F(Q) + w_k \]  

(2)

where \( h \) denotes the vector of state variables (water depth), \( Q \) is the flows from the National Water Model, \( w_k \) is stationary zero-mean white noises, \( F \) represents the prediction model (HAND model), and the subscript "\( k \)" denotes the time step. If an ensemble of \( n \) predicted state variables is available, \( h^i_k \) can be written as:

\[ h^i_k = (h^i_{1k}, . . . , h^i_{nk}) \]  

(3)

where the superscript "\( i \)" represents the \( i \)th forecast ensemble member. Generally, initial condition and inputs are perturbed and ensemble of predicted state variables is generated by running the model for each initial condition. This technique needs to run the model several times for each realization. However, due to computational cost, it is not possible to run the NWM several times to create the ensemble. Instead of perturbing the initial conditions (IC) and inputs, ensemble was created by taking several predicted flows by the NWM up to 4 h (outflow changes by 10% during this period) before the selected events. Outflow was then converted to water depth using rating curves generated by
the HAND. For this study, 10 realizations were created to estimate the error matrix in the model. The average of the ensemble is defined by:

$$\bar{h}_k^f = \frac{1}{n} \sum_{i=1}^{n} h_k^f$$

(4)

Since true states are not known, we estimate them using the average of realizations in the ensemble. Then the error matrix can be estimated by:

$$P_k^f = \frac{1}{n-1} \langle (h_k^f - \bar{h}_k^f)(h_k^f - \bar{h}_k^f)^T \rangle$$

(5)

The error matrix is then used to calculate the Kalman gain matrix by:

$$K_k = P_k^f H_k^T \left( H_k P_k^f H_k^T + R_k \right)^{-1}$$

(6)

where $H$ is the linear transformation which relates the state variables to observations. The updated state vector ($h_k^u$) is taken to be a linear combination of the forecast and the observations. The observations should be treated as random variables to obtain consistent error propagation in the ensemble Kalman filter [29]. Hence, the actual measurements were used as reference and random noise with zero mean and covariance $R$ was added to measurements. The updating equation is given by:

$$h_k^u = h_k^f + K_k (h_k^* - H_k h_k^f)$$

(7)

where $h_k^*$ is the water level observations after adding random noise.

2.5.1. Undersampling

The accuracy of the ensemble Kalman filter is highly dependent on the size of the ensemble. However, due to the complexity of models and computational cost, it is not always possible to generate a large ensemble. If the number of the ensemble is relatively small, it will not be able to accurately estimate the model covariance matrix and system may be undersampled. Inbreeding, filter divergence, and spurious correlation are the main issues caused by undersampling [30]. Covariance inflation and covariance localization are the most common methods used to solve the undersampling problem [31].

2.5.2. Covariance Inflation

Equation (7) is used to inflate the underestimated covariance [32]:

$$h_k^f \leftarrow r (h_k^f - \bar{h}_k^f) + \bar{h}_k^f$$

(8)

where $\leftarrow$ denotes the replacement of a previous value and $r$ is the inflation factor. The optimal inflation factor is related to the ensemble size and may vary between 1.01 and 1.07 [31]. Several methods have been proposed to estimate the inflated forecast and observational error covariance matrices [33]. The adjusted forms of forecast and observational error covariance matrices are $\lambda_k P_k$ and $\mu_k R_k$, respectively. Wu et al. (2013) proposed Equations (9) and (10) to estimate $\lambda_k$ and $\mu_k$ [34].

$$\lambda_k = \frac{\text{Tr} \left( d_k^T H_k P_k H_k^T d_k \right) \text{Tr} \left( R_k^2 \right) - \text{Tr} \left( d_k^T R_k d_k \right) \text{Tr} \left( H_k P_k H_k^T R_k \right) \text{Tr} \left( H_k P_k H_k^T R_k \right)^2}{\text{Tr} \left( H_k P_k H_k^T H_k P_k H_k^T \right) \text{Tr} \left( R_k^2 \right) - \text{Tr} \left( H_k P_k H_k^T R_k \right)^2}$$

(9)

$$\mu_k = \frac{\text{Tr} \left( H_k P_k H_k^T H_k P_k H_k^T \right) \left( d_k^T R_k d_k \right) - \text{Tr} \left( d_k^T H_k P_k H_k^T d_k \right) \text{Tr} \left( H_k P_k H_k^T R_k \right) \left( d_k^T R_k d_k \right) \text{Tr} \left( H_k P_k H_k^T R_k \right)^2}{\text{Tr} \left( H_k P_k H_k^T H_k P_k H_k^T \right) \text{Tr} \left( R_k^2 \right) - \text{Tr} \left( H_k P_k H_k^T R_k \right)^2}$$

(10)
where \( d_k \equiv h^*_{k} - \bar{h}_k \).

### 2.5.3. Covariance Localization

Covariance localization indicated the cutting of the covariance matrix at a specific length [35]. Correlation function suggested by Gaspari and Cohn (1999) was applied to the covariance matrix by using the Schur product to eliminate the spurious correlation [36,37]. The application of covariance inflation and localization methods transform the updating Equation (7) to the one below (Equation (11)).

\[
\begin{align*}
\bar{H}^{*} &= H_k \left( \bar{H}_k \right) + \left[ \rho \left( \lambda_k P^T_k H_k \right) \right] \left[ \rho \left( H_k \lambda_k P^f_k H_k^T \right) + \mu_k R_k \right]^{-1} \left( h^*_{k} - H_k \bar{h}_k \right) \\
\end{align*}
\]

### 3. Results

Streamflow obtained from the NWM for all 85 tributaries were converted to water depth using rating curves generated by the HAND model. Figure 3 shows examples of the rating curves derived from the HAND for two tributaries in the watershed.

![Figure 3. Rating curves created by the Height Above Nearest Drainage (HAND) for (a) Squaw Creek River and (b) Prairie Creek River.](image)

Water level measurements were assimilated into the model for seven bridge sensors selected as training points. Figure 4 shows the water depth at the training locations (a) before data assimilation and (b) after data assimilation, compared with observations from the IFIS. Green triangles are model predictions from the HAND before data assimilation, red squares are updated water depths after data assimilation, and blue dots are water depth observations from the IFIS.

We found that after the assimilation of water depth, the overall error for training locations reduced from 98 cm to 48 cm. We also calculated the error for the testing locations to make sure that the Kalman filter is able to improve the model accuracy for the sites with no observations. It was found that the overall error decreased for testing locations as well (from 89 to 44 cm). Table 2 compares the water depths before and after data assimilation with observations from IFIS for training and testing tributaries. Results show that overall error for both training and testing locations decreased by 48 cm (51%). However, the error has not been improved or error improvement is not significant in some of the sensor locations. This could be due to a small ensemble size and error in the observations (especially error in cross-sections of bridge sensor). The general Root Mean Square Error (RMSE) formula is indicated in Equation (13).

\[
\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left( \text{Predicted}_n - \text{Observation}_n \right)^2}
\]
We found that after the assimilation of water depth, the overall error for training locations reduced from 98 cm to 48 cm. We also calculated the error for the testing locations to make sure that the Kalman filter is able to improve the model accuracy for the sites with no observations. It was found that the overall error decreased for testing locations as well (from 89 to 44 cm). Table 2 compares the water depths before and after data assimilation with observations from IFIS for training and testing tributaries. Results show that overall error for both training and testing locations decreased by 48 cm (51%). However, the error has not been improved or error improvement is not significant in some of the sensor locations. This could be due to a small ensemble size and error in the observations (especially error in cross-sections of bridge sensor). The general Root Mean Square Error (RMSE) formula is indicated in Equation (13).

\[
\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum (\text{observation} - \text{forecast})^2}{n}}
\]

Table 2. Water depth before and after data assimilation compared with water depth measurement from the Iowa Flood Information System (IFIS) for training and testing tributaries.

<table>
<thead>
<tr>
<th>Water Depth (m)</th>
<th>Observation</th>
<th>Before Data Assimilation</th>
<th>After Data Assimilation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training Tributaries</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>0.24</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>1.34</td>
<td>0.37</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>0.72</td>
<td>0.36</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>0.44</td>
<td>1.27</td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>0.69</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>0.49</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>2.39</td>
<td>0.71</td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td>Testing Tributaries</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.30</td>
<td>2.01</td>
<td>1.67</td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td>1.01</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>0.30</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>1.86</td>
<td>1.01</td>
<td>1.99</td>
<td></td>
</tr>
<tr>
<td>Overall RMSE (m)</td>
<td>0.95</td>
<td>0.47</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5 illustrates the time series of flow predicted by the NWM versus flow observation at the USGS station at Squaw Creek at Ames. Water depth observations were assimilated at 6:00, 11:00, and 17:00 on 24 June 2015. After updating the water depth at selected times, the rating curve from the HAND was used to estimate the corresponding flow (green dots in Figure 5). It was found that the overall error between USGS measurements and updated flows from the NWM was reduced by 35% for selected times. Since continental-scale hydrologic models such as the National Water Model are computationally expensive and it is not possible to run them several times to create a larger ensemble, the proposed methodology can be used as an effective method for future studies to update the initial condition of such models.
Acknowledgments: The authors would like to thank the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) and the Office of Water Prediction (OWP) for coordination of Summer Institute. The authors would also like to thank Ibrahim Demir for providing water level observation for state of Iowa.

Author Contributions: Amir Javaheri and Mohammad Nabatian conceived and designed the methodology; Amir Javaheri and Ehsan Omranian prepared the graphs; Amir Javaheri and Ehsan Omranian wrote the paper with the multiple inputs and reviews by Meghna Babbar-Sebens and Seong Jin Noh.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Saksena, S.; Merwade, V. Incorporating the effect of DEM resolution and accuracy for improved flood inundation mapping. J. Hydrol. 2015, 530, 180–194. [CrossRef]


© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).